4

[0038] Where the database is distributed on several servers across a network, the RSD file generation process may need to operate differently, as is described hereinbelow with respect to segmentation.

[0039] In **FIG. 3**, the RSD file **104** is illustrated local to the RSD component **102**. This represents that the RSD file **104** can be generated and stored local to the RSD component **102**, where the RSD component **102** is local to the database **100**. Alternatively, the RSD file **104** can later be moved to the database **100** where the RSD component **102** is remote from the database **100**. All that is important, is that the file **104** is located where it can be readily accessed or provided to a user who needs access thereto.

[0040] Referring now to **FIG. 4**, there is illustrated a general block diagram of an RSD component **400** that further employs a classifier **402** to perform one or more automated functions. The subject invention (e.g., in connection with the RSD file generation process) can employ various artificial intelligence based schemes for carrying out various aspects of the subject invention. For example, a process for determining when the file **104** will be updated by the generation process can be facilitated via an automatic classification system and process. Moreover, where the database **100** is distributed over several locations, and each location has an associated RSD file **104** that represents the distributed database portion at that location, the classifier **402** can be employed to determine which location will be selected for regeneration, in what order the locations will be processed, and when the RSD file **104** will be regenerated.

[0041] A classifier is a function that maps an input attribute vector, x=(x1, x2, x3, x4, xn), to a confidence that the input belongs to a class, that is, f(x)=confidence(class). Such classification can employ a probabilistic and/or statistical-based analysis (e.g., factoring into the analysis utilities and costs) to prognose or infer an action that a user desires to be automatically performed. In the case of database systems, for example, attributes are words or phrases or other data-specific attributes derived from the words (e.g., database tables, the presence of key terms), and the classes are categories or areas of interest (e.g., levels of priorities).

[0042] For example, a support vector machine (SVM) classifier can be employed. An SVM operates by finding a hypersurface in the space of possible inputs. This hypersurface will attempt to split the triggering criteria from the non-triggering events. Intuitively, this makes the classification correct for testing data that is near, but not identical to the training data. Other directed and undirected model classification approaches include, e.g., naïve Bayes, Bayesian networks, decision trees, and probabilistic classification models providing different patterns of independence can be employed. Classification as used herein also is inclusive of statistical regression that is utilized to develop models of priority.

[0043] As will be readily appreciated from the subject specification, the subject invention can employ classifiers that are explicitly trained (e.g., via a generic training data) as well as implicitly trained (e.g., via observing user behavior, receiving extrinsic information). For example, SVM's are configured via a learning or training phase within a classifier constructor and feature selection module. Thus, the classifier(s) can be used to automatically determine according to a predetermined criteria when the database RSD file

is generated, when it is regenerated (i.e., updated), which RSD file of a distributed database is to be regenerated, etc. The criteria can include, but is not limited to, the amount of data to represented in the RSD file **104**, the amount of data to be mapped (in a mapping operation), the type of data, the importance of the data, and the level of confidence requested for generating the RSD file **104**.

[0044] Referring now to **FIG. 5**, there is illustrated a block diagram of a system where the RSD component **102** (or **400**) is located remote to the relational database from at least one relational database **100**. The database **100** and the RSD component **104** are disposed in communication on a network **502**, e.g., the Internet. Thus, the RSD component **104** can be employed to remotely connect to the database **100** and perform generation of the RSD file that represents the structure and data thereof. The RSD file can then be stored local to the database **100** for access and use for various purposes, as described hereinabove.

[0045] There is provided a second relational database **504** disposed on the network **502** such that the second database **504** can be a standalone database, or distributed in combination with the database **100**. The disclosed RSD schema supports distributed database environments. That is, an associated RSD sub-file can be generated for each distributed portion of the overall database. The user can then retrieve one or more sub-files depending upon the data that he or she wishes to access. It is to be appreciated that the RSD file, whether the file is one that represents the total database, or a sub-file that represents only a portion thereof in a distributed environment, may be generated depending upon the access rights of the user. For example, if the user has low-level access rights, and requests access to the database, the tool can automatically generate an RSD file that only represents that portion of the database that would be accessible to the user had the user been connected to the database under normal conditions. Similarly, if the user has high-level rights (i.e., global or administrative-type access), the tool can automatically access the user profile to determine that the RSD file to be generated will represent all aspects of the database that the user could access under normal conditions.

[0046] In the context of a classifier operating in conjunction with the disclosed architecture, the classifier can be employed to automatically learn and anticipate which users should have an RSD file generated for ready access, and how much of the database to use for generating the file. For example, as the database **100** is repeatedly accessed over time, the classifier can track the access behavior of the particular users. Thus, when the database activities, or other triggering criteria trigger auto-regeneration of the RSD file, the classifier can auto-regenerate an RSD file for each of the anticipated users. If a high-level employee routinely accesses the database for financial information on Friday evening at 9 PM, the classifier can learn this and have the most recent RSD schema representation ready in file format at that time for access by the employee.

[0047] In another scenario, if the classifier detects that Monday morning is a very active time for database accessing by employees, the classifier can learn this and prepare more recent RSD file regenerations at the early times of the morning. In this same context, if the classifier is programmed to include a fixed number of people in a process,